DOCUMENT RESUME

ED 233 088 TM 832 005

TITLE	Performance Assessment Methodology. Resources in Performance Assessment
INSTITUTION	Northwest Regional Educational Lab., Portland, OR. Center for Performance Assessment.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	[83]
CONTRACT	400-83-0005
NOTE	6p.
PUB TYPE	Reference Materials - Bibliographies (131)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Achievement Tests; Administrator Evaluation; Annotated Bibliographies; Assessment Centers (Personnel); Critical Incidents Method; *Evaluation Methods; Generalizability Theory; *Performance; Performance Tests; Personnel Evaluation; Student Evaluation; Test Construction; *Testing
IDENTIFIERS	PF Project

ABSTRACT

This annotated bibliography contains nine items addressing assessment methodology. The titles are: "Performance and Product Evaluation"; "The Critical Incident Technique"; "Constructing Achievement Tests"; "Applying the Assessment Center Method"; "Performance Assessment in Education and Training: Alternative Techniques"; "Watching Students Grow: A Teacher's Guide to Observational Assessment in the Classroom"; "Behavioral Assessment" (a journal); "Generalizability Theory: A Review"; and "Behavioral Assessment" (a book). These entries include books, published articles and a professional journal. (JW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. Margaret Rogers

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED233088

Center for Performance Assessment

RESOURCES IN PERFORMANCE ASSESSMENT

Northwest Regional Educational Laboratory

measuring achievement via behavior & product ratings

Performance Assessment Methodology

TM 832 005

Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue • Portland, Oregon 97204

Center for Performance Assessment

RESOURCES IN PERFORMANCE ASSESSMENT

Northwest Regional Educational Laboratory

measuring achievement via *behavior & product ratings*

Performance Assessment Methodology

Covering the Basics

FITZPATRICK, ROBERT AND EDWARD J. MORRISON. "Performance and product evaluation." In R.L. Thorndike (Ed.) *Educational Measurement*. Washington, D.C.: American Council on Education, 1971, 237-270.

During each of the last four decades, performance assessment practice has been strongly influenced by several prominent measurement scholars. In 1951, for example, Ryans and Fredericksen presented their views on performance assessment in E.F. Linquist's (Ed.) *Educational Measurement* (Washington, D.C.: American Council on Education). Then Glaser and Klaus in 1962 provided a major discussion of the topic in *Psychological Principles of System Development* by R.M. Gagne (Ed.), (New York, NY: Holt, Rinehard and Winston). Almost ten years later, in 1971, the Fitzpatrick and Morrison chapter appeared. These were followed in 1982 by the Priestly book (reviewed below) on alternative methods of using performance assessment. Those interested in the development of thought on performance assessment will find it described in this series of publications.

The discussion of performance assessment by Fitzpatrick and Morrison (1971) addresses the topic from some unique perspectives. For example, the authors give far more attention to the concept of simulation, the application of high technology in performance assessment, the basic concepts of fidelity, cost, and the essentials of good simulations. Their discussion of the role of simulation in performance assessment remains very relevant in the 1980s.

Second, the authors emphasize the use of situation tests such as in-basket tests, work sample tests, games, contests and diagnostic problem-solving tests in which examinees engage in some real life tasks. Unlike previous discussions of the topic, this one focuses extensively on applications of performance assessment in many occupational and personal contexts.

However, like earlier discussions, this overview of performance assessment includes guidelines for developing performance tests, which cover test specifications, stimulus conditions, response modes, conditions for appropriate test administration and scoring, and guidelines for test use. This chapter by Fitzpatrick and Morrison is relevant for assessment scholars and practitioners alike.

FLANAGAN, JOHN C. "The critical incident technique." *Psychological Bulletin*, 1954, 51(4), 327-358.

A key to successful performance assessment is to focus assessment on the essential characteristics of effective performance. But how do we determine important criteria of performance? Several methods are available. One may observe the performance of qualified experts, or ask these skilled professionals to describe keys to success.

John Flanagan developed yet another effective means of identifying performance criteria—the critical incident technique. His 1954 description of that method is a classic description of the systematic assessment of human performance.

To identify the key attributes of effective performance, Flanagan recommends five procedures: (1) observe the activity's purpose or aim; (2) specify details of the observation, including the number and characteristics of observers and the person(s) and behaviors to be observed; (3) collect data via interviews, group discussion, questionnaires and/or existing records; (4) analyze observational data; and (5) interpret and report results.

Flanagan describes many job-related uses of the critical incident technique, including methods to establish performance criteria, design measures of proficiency, train, select and classify employees, design jobs, clarify operational procedures, and counsel employees.

Though the critical incident technique originated in industry, it has important implications for educational assessment. An analysis of critical incidents is one of the best ways to determine characteristics of successful performance. For example, when a student performs well in an oral presentation, what key behaviors reflect that success? When performance is inadequate, what is missing? These questions help identify the performance criteria that should guide performance ratings. The more explicitly teachers document those behaviors, the easier it will be to train others to identify critical dimensions of student performance.

RESOURCES IN PERFORMANCE ASSESSMENT presents abstracts of selected publications on important aspects of performance assessment. The selections offer educators easy access to some of the most recent and useful publications available on various important assessment topics. Because space in the bibliography is limited, these references should be viewed as only a representative sample of relevant resources on this topic.

GRONLUND, NORMAN E. *Constructing Achievement Tests* (3rd Ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc., 1982, 148 pp.

Chapter 6 of this book, entitled "Constructing Performance Tests," offers one of the most useful presentations on performance assessment available in an introductory measurement textbook. Brief and to the point, this chapter gives teachers and others an overview of the various types of performance assessments and simple procedural guidelines for developing them.

Gronlund's chapter begins with a clear description of the far-ranging applicability of performance assessment in education: "Performance tests are concerned with skill outcomes. Skill in using processes and procedures is a desired outcome in many academic courses. For example, science courses are typically concerned with laboratory skills, mathematics courses are concerned with practical problem-solving skills, . . ."

Four types of performance tests are recommended for measuring these skill outcomes: (1) paper and pencil performance tests, (2) identification tests, (3) simulated performance, and (4) work samples. Paper and pencil performance tests assess students' ability to apply skills by having them solve written problems. Identification tests – sometimes referred to as indirect measures of performance – ask students to identify the tools used in industrial education, science, language or the arts. The assessor infers that if tools are identified, they can also be properly used. Simulated and work sample performance tests measure an individual's ability to carry out procedures or produce products required in school, personal life or work.

Gronlund also outlines four steps for constructing a performance test. Step one specifies the performance outcomes to be measured and describes acceptable performance. Step two involves selection of an appropriate degree of realism for the test. The author describes several practical factors that affect realism.

The third step includes a clear specification of the conditions within which skills are to be demonstrated. And the final step calls for preparation of the observational form for evaluating performance.

Gronlund's text, *Measurement and Evaluation in Teaching* (New York, NY: MacMillan Publishing Co., 1976), also presents a comprehensive discussion of this topic. Both are excellent starting points for studying performance assessment.

MOSES, JOSEPH L. AND WILLIAM C. BYHAM. *Applying the Assessment Center Method.* New York, NY: Pergamon Press, 1977, 310 pp.

An assessment center, as described by Moses and Byham, is both a place and a process. As such, the center includes a collection of performance tests, each designed to provide experts with an opportunity to rate examinee performance in critical aspects of job functioning. The assessment center allows qualified judges to observe and evaluate many dimensions of performance and then review, interpret and provide feedback to examinees. Developed for use in personnel selection, these centers offer job candidates an opportunity to demonstrate job-relevant skills in simulated performance assessment situations.

During the 1960s and 1970s, assessment centers were used for identifying employees' management, sales and technical potential. Though assessment centers have their roots and major applications in business and industry, they also have potential for profitable educational use. Stanford Graduate School of Business, Alverno College and Nova University have used assessment centers to measure students' achievement, and organizations such as the American Association of School Adminstrators have used this approach to help administrators identify school management skills in need of development.

This book is both practical and comprehensive. Principles of center development and use are discussed in a series of chapters written by practitioners who have designed a variety of performance assessments. The authors also discuss issues in the evaluation and future development of assessment centers.

Appended to the report is an instructive set of "Standards and Ethical Considerations for Assessment Center Operations" developed by a task force and approved by professionals in the field. These brief standards define assessment centers and offer guidelines for organizational use, training assessors, preserving the rights of examinees, using data and ensuring assessment quality. Any assessor designing and using performance tests – whether in educational or business contexts – would do well to refer to these directives.

Those interested in recent developments related to assessment centers since the 1977 publication of Moses and Byham's book are referred to the *Journal of Assessment Center Technology* published by Assessment Designs, Inc., (ADI Court, 601 N. Ferncreek Avenue, Orlando, Florida 32803).

The *Center for Performance Assessment*, a research and dissemination project funded by the National Institute of Education, serves educators by conducting research on performance assessment – the observation and rating of student behavior and/or products – and by disseminating resource information on this assessment method. Established at the Northwest Regional Educational Laboratory in 1973, the Center develops bibliographies, monographs and a regular newsletter entitled CAP-TRENDS, and conducts workshops. It also provides technical consultation to educators on the development and use of performance assessment to measure students' skills. For information on publications and services provided by the Center, please contact:

Center for Performance Assessment
Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, Oregon 97204
248-6800 x352 (in Oregon)
1-800-547-6339 (outside Oregon)

This bibliography was developed by the Center for Performance Assessment under contract #400-83-0005 with the National Institute of Education (NIE), Department of Education. The opinions expressed do not necessarily reflect the position or the policy of NIE and no endorsement by NIE should be inferred.

PRIESTLY, MICHAEL. *Performance Assessment in Education and Training: Alternative Techniques.* Englewood Cliffs, NJ: Educational Technology Publications, 1982, 263 pp.

For specialists in educational measurement, this book should be a basic reference. Priestly describes 25 different types of performance assessments in terms of their form, uses, advantages, disadvantages and, most importantly, steps in test development. He includes many concrete illustrations of assessments and addresses keys to successful administration and scoring. This represents the most comprehensive treatment of the topic to date.

Included among these types of assessment are actual performance tests, including work samples, identification tests, and supervisor, peer and self-ratings. Simulations are described, covering job simulations, written simulations and management simulations, among others. Observational assessments, such as checklists, rating scales and anecdotal records are summarized, as are oral assessments, paper and pencil assessments and personnel records. Priestly also discusses the use of assessment centers and performance appraisal systems.

Priestly's range of alternative methods plus Gronlund's description of subject matter areas (described previously) clearly indicate the potential of performance assessment in a wide range of school settings.

STIGGINS, RICHARD J. *Watching Students Grow: A Teacher's Guide to Observational Assessment in the Classroom.* Portland, OR: Center for Performance Assessment, Northwest Regional Educational Laboratory, 1982, 37 pp.

This publication offers teachers specific, practical guidelines for using performance assessment to measure student behavior and/or products and specifies procedures to ensure test quality. The author contends that systematic test design and careful quality control make performance assessments as objective, and therefore as useful, as any other form of classroom assessment.

To maximize the use of quality performance assessments, the author outlines a step-by-step test development sequence. Within each step, the teacher makes a series of specific planning decisions and considers various factors described in the text for selecting alternatives.

In step one, the teacher describes the assessment situation, covering the specific reason for testing, decision makers who will use test results, and the skills and/or knowledge to be demonstrated.

Next, the stimulus event is defined in terms of the task(s) students will be asked to perform. These may involve either naturally occurring events or simulations, but must include enough activities to ensure confidence in results.

In step three, teachers must describe the student's response to be evaluated, stipulating (1) whether a process (behavior, procedure) or product (result of doing) is to be rated, (2) what criteria will be used to judge performance, and (3) whether or not students are to be informed of the performance evaluation.

And in the final step, teachers plan rating procedures by selecting scoring methods and evaluators (teacher, another expert, students, self or peers), and by determining whether results will be interpreted in terms of norms or preset minimum standards.

After guiding the teachers through these planning steps, the author specifies directions for ensuring quality assessment. These guidelines promote clear testing purposes, effective communication about assessment, maximum objectivity and economy of test use.

As a concluding point, Stiggins differentiates between structured, preplanned tests of performance and spontaneous observations of classroom behavior, and recommends techniques for maximizing the quality of both types of performance assessments.

Beyond the Basics

Behavioral Assessment. Journal of the Association for the Advancement of Behavior Therapy. New York, NY: Pergamon Press.

A good source of current information for the measurement specialist on methodological developments in performance assessment is the journal *Behavioral Assessment*. Though most articles focus on performance assessment in clinical psychology, nearly every issue contains at least one article of interest to educational measurement specialists, and no other single journal provides more up-to-date technical information on performance assessment.

To illustrate, here are three educationally useful articles that appeared recently:

In "Detecting bias in observational data" (Volume 2, 1980), the author, House, recommends methods for analyzing data to detect systematic error. When two observers record observations of dichotomous variables, rater agreement can be summarized *and* cell frequencies compared to determine if one rater is consistently higher or lower than the other.

Haynes and Horn review some 30 studies on "Reactivity in behavioral observation" (Volume 4, 1982). Reactive effects occur when the process of evaluating performance results in a permanent or temporary change in a student's behavior. For example, motivational or anxiety factors from the assessment may bring about atypical performance. The authors discuss ways to detect, mediate and overcome reactive effects. To minimize effects, they suggest: (1) using participant observers, (2) evaluating products rather than behaviors, (3) using covert observations or recording devices (e.g., video), (5) instructing students to "act naturally," and (6) delaying observations to allow time for reactive effects to dissipate.

Hart's article outlines systematic procedures for "Assessing spontaneous speech" (Volume 5, 1983). Claiming that "all language training programs need to include assessment of how children in training actually use language outside of training," the author offers a method for the longhand recording of spontaneous speech in instructional settings. These assessment methods may be of value to all language arts educators.

These are only three of numerous educationally relevant articles in *Behavioral Assessment*. This journal provides an excellent, up-to-date record of practical and theoretical developments in the field, and is an important sequel to Haynes and Wilson's *Behavioral Assessment* volume, which summarized research and development through 1978.

BRENNAN, ROBERT L. AND MICHAEL T. KANE. "Generalizability theory: A review." *New Directions for Testing and Measurement*. San Francisco, CA: Jossey-Bass Publishers, 1979, 33-51.

To evaluate performance, we sample from all possible instances of that performance to predict how well the student will perform in other instances. Like all tests, performance tests sample a domain of behavior. The test developer's goal is to sample behavior accurately so confident generalizations can be made.

However, in most performance assessments, various problems can occur that result in undependable assessments. For example, performance exercises may not represent a typical activity, causing us to make inappropriate generalizations to the intended behavior domain. Further, performance rating procedures may be inaccurate or raters inadequately trained. As a result, measurement errors may occur.

Generalizability analysis provides a means of both estimating the dependability of ratings and determining the source and size of measurement errors. In this chapter, Brennan and Kane explain how this is accomplished. They cover basic concepts, principles, procedures, and uses of generalizability theory. The authors assume readers have some familiarity with classical test theory and analysis of variance.

Those concerned with large-scale performance assessments, such as statewide writing assessments using writing samples, for example, can address key issues of test score reliability only by applying generalizability analysis. Brennan and Kane describe specific models and methods for partitioning variance and constructing the variance ratios needed to fully explain sources of unreliability in performance ratings. This method also allows test developers to generalize beyond existing data and determine the impact of potential measurement errors in test exercises and/or scoring on score reliability in performance assessments.

HAYNES, STEPHEN N. AND C. CHRISMAN WILSON. *Behavioral Assessment*. San Francisco, CA: Jossey-Bass Publishers, 1979, 526 pp.

Behavioral assessment—based on the observation of actual behavior—is as important to the clinical psychologist as performance assessment is to the classroom teacher. Haynes and Wilson illustrate this by referencing over 70 journal articles addressing behavioral assessment in educational contexts. Specialists in educational measurement will find the description of developments in behavioral assessment and their educational applications both interesting and useful.

In discussing issues in behavioral assessment, the authors describe a concise set of criteria for evaluating behavioral (or performance) assessment. Evaluation criteria include: (1) utility for the intended population and purpose; (2) sensitivity to real changes in student performance; (3) reliability or consistency of assessment (including internal consistency, consistency over time, and interobserver agreement); (4) validity of the assessment (content, criterion-related, construct); and (5) changes in the target behavior resulting from assessment. Each evaluation criterion is illustrated with examples of assessments.

The publication also includes reference to current research literature on such recent methodological advances as behavioral coding systems, strategies for assessing interobserver agreement, and advances in participant observation; observation in natural environments, including schools; observations of child behavior in structured learning environments; and use of self-monitoring, behavioral questionnaires and behavioral interviews.

All 71 references on assessments in school settings appeared in research journals between 1977 and 1979—an average of well over 20 references per year. This indicates the growing role behavioral assessment played in school research during this brief period. For references on the use of behavioral assessment in educational research after 1979, refer to the 1980-82 issues of the journal, *Behavioral Assessment*, described previously.